



Non-linear retinal processing supports invariance during fixational eye movements



Garrett Greene^{a,b,*}, Tim Gollisch^{b,c,d}, Thomas Wachtler^{a,b}

^aDepartment Biology II, Ludwig-Maximilians-Universität München, Germany

^bBernstein Center for Computational Neuroscience Munich, Germany

^cDepartment of Ophthalmology, University Medical Centre, Göttingen, Germany

^dBernstein Center for Computational Neuroscience Göttingen, Germany

ARTICLE INFO

Article history:

Received 30 July 2014

Received in revised form 1 October 2015

Accepted 18 October 2015

Available online 12 November 2015

Keywords:

Retinal ganglion cell
Fixational eye movement
Ouchi illusion
Parasol
Y cell
Motion
Stabilisation
Gaze-invariance
Invariance

ABSTRACT

Fixational eye movements can rapidly shift the retinal image, but typically remain unnoticed. We identify and simulate a model mechanism for the suppression of erroneous motion signals under fixational eye movements. This mechanism exploits the non-linearities common to many classes of large retinal ganglion cells in the mammalian retina, and negates the need for extra-retinal signals or explicit gaze information. When tested using natural images undergoing simulated fixational eye movements, our model successfully distinguishes “real world” motion from retinal motion induced by eye movements. In addition, this model suggests a possible explanation for several fixational eye movement related visual illusions such as the Ouchi–Spillmann and “Out-of-focus” illusions.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Fixational eye movements and retinal motion

A constant challenge for the visual system is to resolve a single stable percept from an often ambiguous or heavily confounded retinal signal.

This is especially true in the case of motion perception, where the retinal signal is highly confounded by the self motion of the observer, and in particular by constant rapid eye movements. Numerous studies have shown that the human eye is in almost constant motion, even during periods of attempted fixation, and indeed, it is known that small eye movements play an essential role in visual processing (Ahissar & Arieli, 2012). However, such motions can also have the effect of creating spurious image motion on the retina, which can be difficult to distinguish from behaviourally relevant real-world motion. Fixational eye movements (FEM), including drifts, tremors and micro-saccades can rapidly and unpredictably shift the retinal image across tens or hundreds of

photoreceptors. These movements are easily large enough to excite motion-sensitive retinal ganglion cells, and thus evoke a motion percept (Steinman, Haddad, Skavenski, & Wyman, 1973). The fact that we perceive the world as remaining stable during fixational epochs suggests that there must be a mechanism capable of suppressing erroneous motion signals.

In the case of self motion of the head or body, or large visually guided saccades, it is known that the visual cortex receives input from the vestibular system, as well as feedback from motor pathways containing information about body position and gaze direction, which can be used to rectify the visual percept with respect to these motions (Gregory, 1958; Perrone & Krauzlis, 2008). However, it remains unclear whether such extra-retinal signals are sufficient to stabilise motion percepts in the case of fixational eye movements (Martinez-Conde, Otero-Millan, & Macknik, 2013). Furthermore, whereas it is known that visual processing in the cortex is broadly suppressed during visually guided saccades, small fixational eye movements are believed to be an essential component of active vision, making visual suppression during such movements highly unlikely and undesirable. While recent studies have provided strong evidence for extra-retinal mechanisms of micro-saccadic suppression (Hafed & Krauzlis, 2010; Hafed, 2011), it is also known that micro-saccades can enhance activity

* Corresponding author at: Department Biology II, Ludwig-Maximilians-Universität München, Germany.

E-mail address: greene@bio.lmu.de (G. Greene).

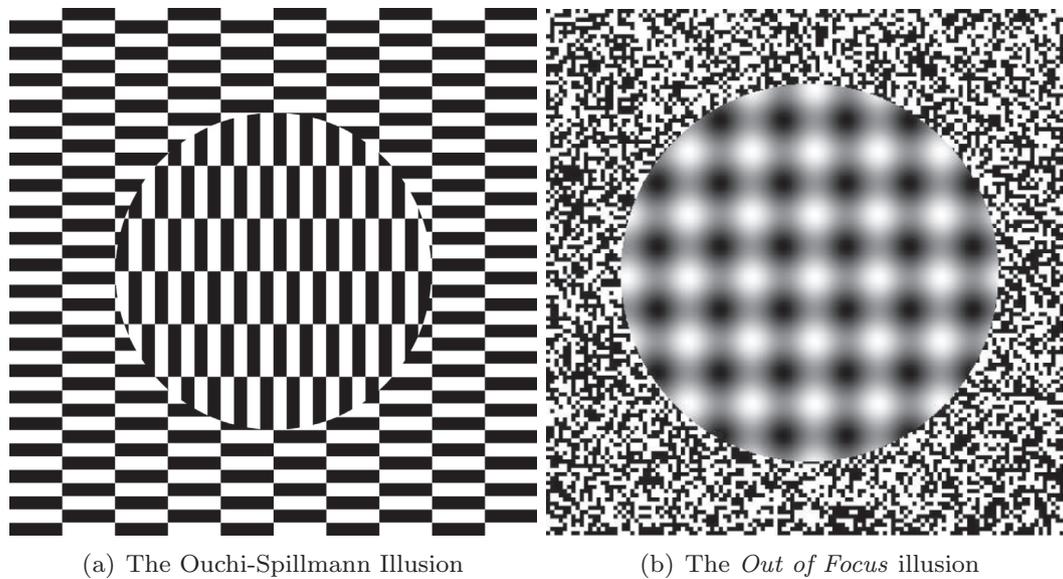


Fig. 1. The Ouchi–Spillmann illusion (left) and Kitaoka’s “Out of Focus” illusion (right) (Kitaoka, 2001; Ouchi, 1977; Spillmann & Werner, 2012). When viewed from the correct distance both images produce an illusory jittery motion of the central region relative to the background. In both cases the motion effect depends on the sharp boundary between areas of high and low spatial frequency, and can be induced by fixational eye movements. The effect is weakened when the viewer makes a concentrated effort to keep their eyes still, thus reducing the frequency of large eye movements.

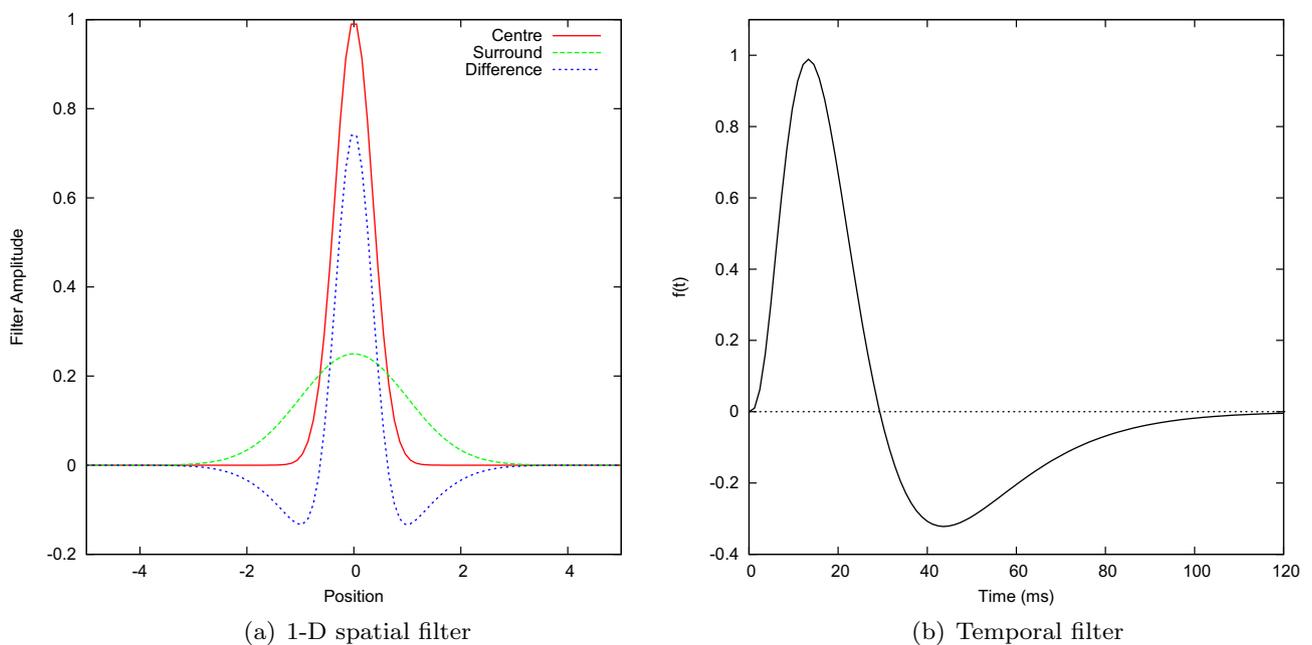


Fig. 2. Left: a 1-D difference of Gaussians spatial filter (Eq. (1)) displaying the well-known “Mexican Hat” profile. Right: an example bi-phasic temporal filter (Eq. (2)) with time constants $\tau_1 = 5$ ms and $\tau_2 = 10$ ms.

in early visual cortex (Martinez-Conde, Macknik, & Hubel, 2000; Meirovithz, Ayzenshtat, Werner-Reiss, Shamir, & Slovlin, 2011). Meanwhile, there is no known suppressive mechanism for motions due to tremors or linear drift, which can be well above motion detection thresholds. In addition, it is known that the suppression of motion percepts due to FEM fails under certain stimulus conditions, suggesting that the suppression mechanism is of retinal origin (Murakami & Cavanagh, 1998; Otero-Millan, Macknik, & Martinez-Conde, 2012; Poletti, Listorti, & Rucci, 2010). Therefore, it is believed that the visual system must be capable of extracting information about fixational movements from the retinal activity alone, and thus distinguishing real-world motion from

eye-motion through inference on the statistics of the retinal signal. This notion is supported by the existence of several visual motion illusions which appear to be induced by fixational eye movements. In the case of both the Ouchi–Spillmann illusion, and Kitaoka’s “Out of focus” illusion, observers perceive transient motion or jitter within stationary images (Martinez-Conde, 2006; Otero-Millan et al., 2012). The perception of illusory motion in both cases is strongly correlated with the frequency of fixational eye movements (Murakami, Kitaoka, & Ashida, 2006). These illusions thus provide some insight into the mechanism of retinal motion processing under FEM. Here we propose a model in which motion contrast is used to distinguish local motion, which is interpreted as

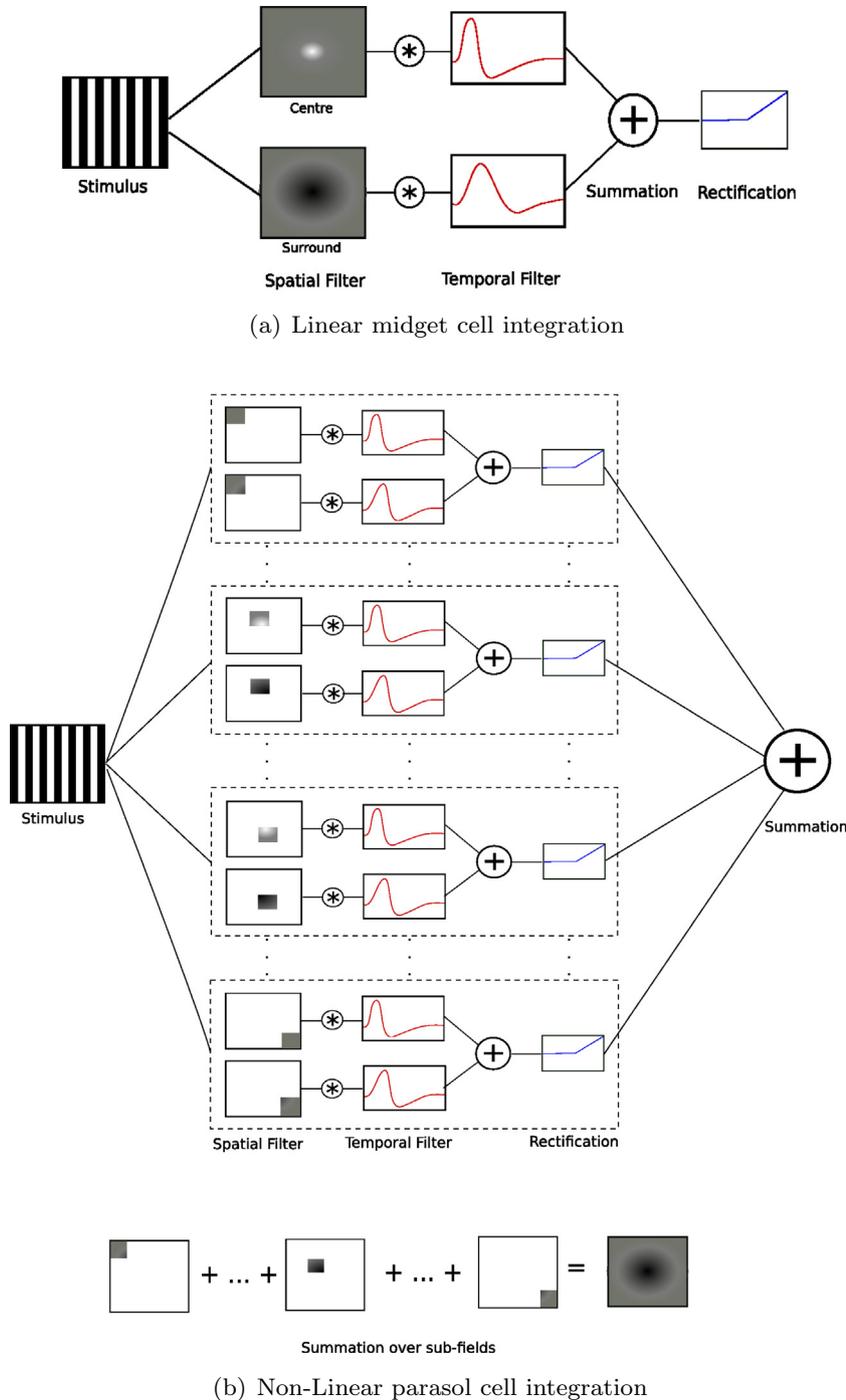


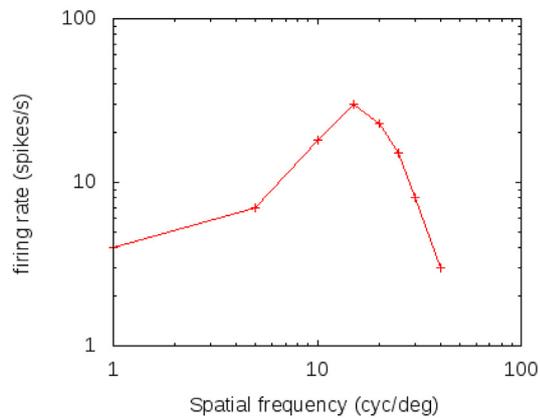
Fig. 3. Schematics of linear (top) and non-linear (bottom) RGC models. The linear cell sums contributions across the entire RF before rectification. In the non-linear model, individual contributions from sub-fields are rectified before summation.

real-world motion, from coherent global motion of the retinal image, such as that induced by eye-movements. It exploits the non-linear response properties of certain classes of large retinal ganglion cells which are common in the vertebrate retina. The model straightforwardly leads to a stabilised motion percept under simulated FEM and in addition reproduces the illusory motion effect of the Ouchi illusion and other related motion illusions.

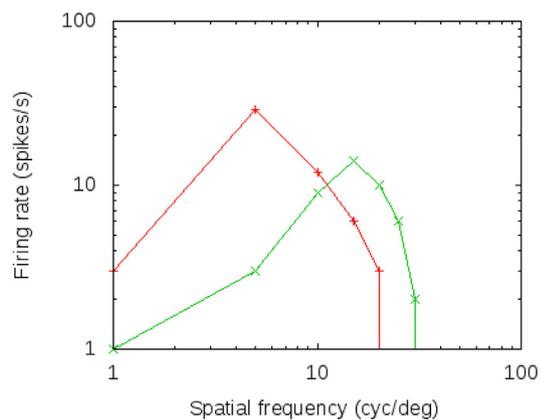
1.2. Models for cancellation of retinal motion

The notion of cancelling retinal motion signals caused by FEM is not an entirely new one. Numerous studies have addressed the

perceptual effects of FEM, and suggested models which can compensate for visual blurring resulting from constant eye motion. Pitkow et al. and Burak et al. present elegant Bayesian decoder models which can correct for retinal smearing of the image and allow high acuity visual discrimination, while Ahissar et al. propose a model in which FEM allow for visual hyperacuity through a latency coding scheme linked to eye motion (Ahissar & Arieli, 2012; Burak, Rokni, Meister, & Sompolinsky, 2010; Pitkow, Sompolinsky, & Meister, 2007). However, while such models can compensate for the blurring due to retinal motion, or indeed incorporate motion into a sampling strategy, they do not address the additional problem posed by FEM, namely that of how to



(a) Spatial frequency response of model midget cell



(b) Model parasol cell spatial frequency response showing Y-Cell signature. First (red) and second (green) harmonic response amplitudes to sinusoidally reversed gratings are plotted as a function of grating spatial frequency. Grating reversal frequency was $2Hz$. At higher spatial frequencies the second harmonic is dominant.

Fig. 4. Responses of midget and parasol cells to reversing gratings as a function of grating spatial frequency.

distinguish retinal motion from real-world motion. While the bayesian estimator model allows for the corrected estimation of object position in retinal coordinates, it cannot determine whether the motion is that of the object or of the observer, and so does not distinguish eye motion from real-world target motion.

The question of motion correction has been more directly addressed by [Murakami and Cavanagh \(1998\)](#). In their experiment, subjects who had been adapted to a patch of dynamic random noise reported a coherent jitter effect when viewing static stimuli, believed to be directly linked to FEM. They provide compelling evidence that the visual system employs some method of differential motion processing based on purely retinal signals, and propose a model in which the estimated eye motion is subtracted from the retinal motion signal. Murakami subsequently found that horizontal and vertical motion thresholds correlate with FEM velocity ([Murakami, 2004](#)), providing further evidence for the hypothesis of differential motion processing. However, as reported by [Tong et al.](#), this does not hold true for torsional eye movements ([Tong, Lien, Cisarik, & Bedell, 2008](#)). In addition, while offering an explanation of their own results, the subtractive model of Murakami and Cavanagh does not explain static motion illusions such as the Ouchi illusion, which produce similar jitter effects without the need for adaptation. Indeed, within this framework, Murakami

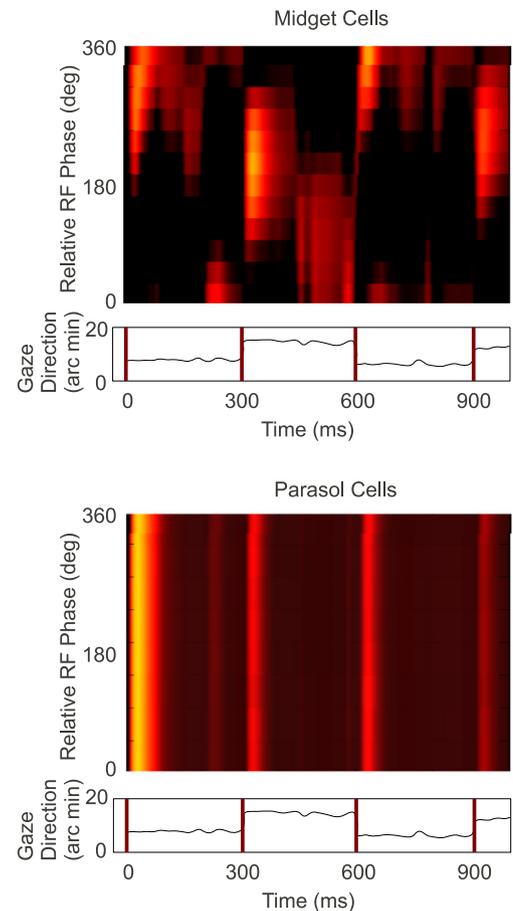


Fig. 5. Responses of midget and parasol cells to grating stimulus under simulated FEM. The response of midget cells is strongly modulated with the phase of the grating (Y-axis) as expected from the centre-surround receptive field. By contrast parasol cells respond in a phase-invariant manner, and display a strong transient response to large motions such as microsaccades -indicated by red lines.

et al. present evidence of illusory motion effects attributable to FEM which are not explained by this model ([Murakami et al., 2006](#)).

The model proposed here differs from that of Murakami in several ways. In particular, our model does not require the computation of a retinal velocity, but rather cancels motion signals locally within sub-regions of the retina. This cancellation mechanism arises naturally from the non-linearity of parasol retinal ganglion cells, and allows invariance to be achieved more rapidly, and at lower cost than other methods which rely on lateral interactions between motion detecting units. In addition, this model depends explicitly on the spatial frequency response properties of retinal cells, and thus predicts the failure of the motion cancellation in the case of certain images with non-uniform spatial frequency statistics, such as seen in the Ouchi illusion.

1.3. Non-linear retinal ganglion cells

Classical models of retinal coding describe the stimulus response of retinal ganglion cells (RGCs) in terms of the linear receptive field ([Enroth-Cugell & Robson, 1966](#); [Enroth-Cugell, Robson, Schweitzer-Tong, & Watson, 1983](#); [Kuffler et al., 1953](#); [Rodieck, 1965](#)). In this context, the Receptive Field (RF), represents the stimulus to which the neuron in question is expected to respond most strongly. Equivalently, the RF can be used to predict neuronal firing, and is defined as the linear filter which, when

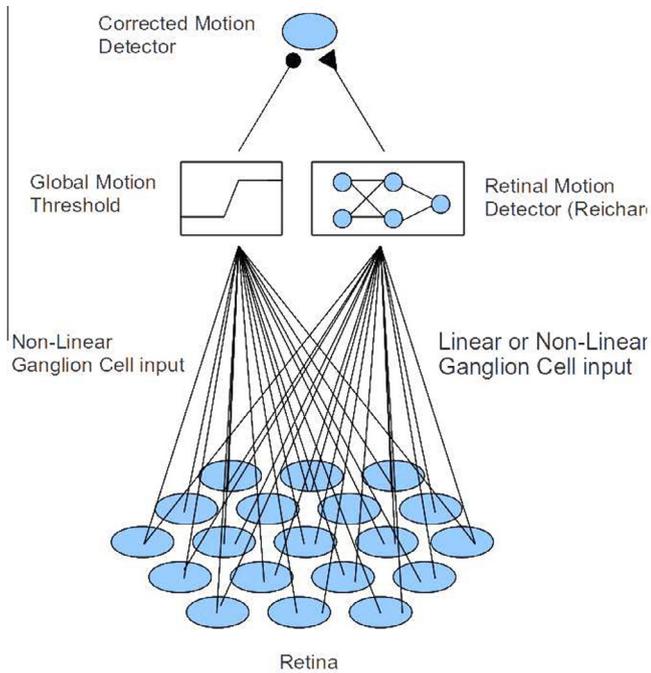


Fig. 6. Schematic of motion detection and cancellation circuit. Motion is computed locally by a population of motion detector circuits, each with distinct direction and velocity tuning. The detected motion vector is determined by a weighted sum over the motion vectors associated with each detector. This motion signal is then gated by the output of a wide-RF global motion detector, which detects population responses in the parasol cell population.

convolved with the stimulus, gives the best firing rate prediction. In the case of RGCs, the spatial component of the RF commonly takes the form of the well-known centre-surround filter, consisting of concentric and opponent centre and surround regions. Cells can be classified as either On-type (with bright centre and darker surround) or Off-type (dark centre, bright surround).

In signal processing terms, this structure has a simple functional interpretation as a contrast-enhancing filter, which reduces spatial redundancy in the signal, leading to the common assumption that the role of centre-surround RGCs is simply to encode spatial contrast for visual scenes, with On- and Off- types respectively encoding positive and negative contrasts (Marr & Hildreth, 1980; Wilson & Giese, 1977).

However, there exists a large class of centre-surround RGCs, such as the Y-Cells of the cat, or parasol cells of the macaque retina, which display complex responses that cannot be explained by the linear receptive field model. These include transient responses to motion and saccades, as well as phase invariant responses to grating stimuli (Baccus, Ólveczky, Manu, & Meister, 2008; Bölinger & Gollisch, 2012; Enroth-Cugell & Robson, 1966; Gollisch & Meister, 2010; Hochstein & Shapley, 1976b; Kremers, Lee, Pokorný, & Smith, 1993; Münch et al., 2009). Here we employ a model of RGC processing which explains these non-linear effects as the result of summation over independent sub-fields in the RF (Lee, Kremers, & Yeh, 1998; Victor, Shapley, & Knight, 1977). An interesting consequence of these non-linearities is that such cells, despite having On- or Off- receptive fields, can in fact display both On- and Off- responses to onset of natural image or grating stimuli, invariant to the local phase of the stimulus. This is closely related to the frequency doubling effect, characteristic of Y-type cells, which display a second harmonic or frequency-doubled response to periodically changing stimuli (Hochstein & Shapley, 1976a; Shapley & Perry, 1986; Victor & Shapley, 1979). As a result, such cells may display an invariant transient response to stimulus motion. We show here that this property can easily lead to a

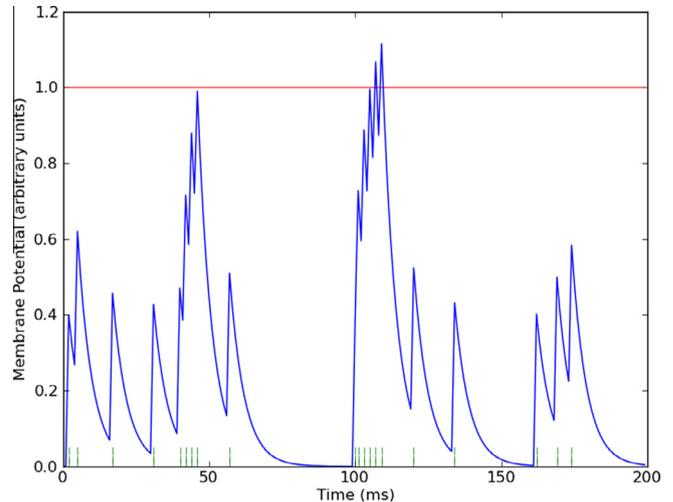


Fig. 7. Membrane potential of a motion threshold neuron. Green pulses indicate arrival time of incoming action potentials. The cell requires a minimum number of near co-incident spikes to exceed threshold. Threshold units are tuned so that on average 50% of afferent parasol cells must fire within a 10 ms window to reach threshold.

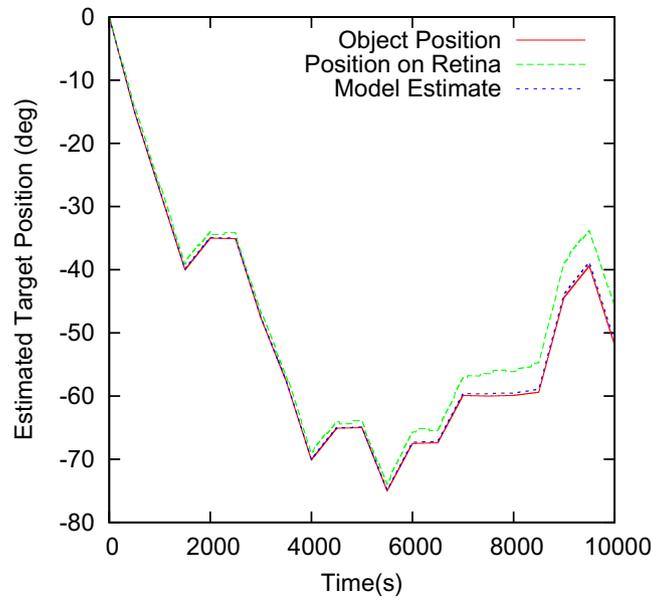
saccade-detection mechanism, whereby the simultaneous activity of a population of such cells signals a global motion of the image.

1.4. Ouchi illusion and Kitaoka's "Out of focus" illusion

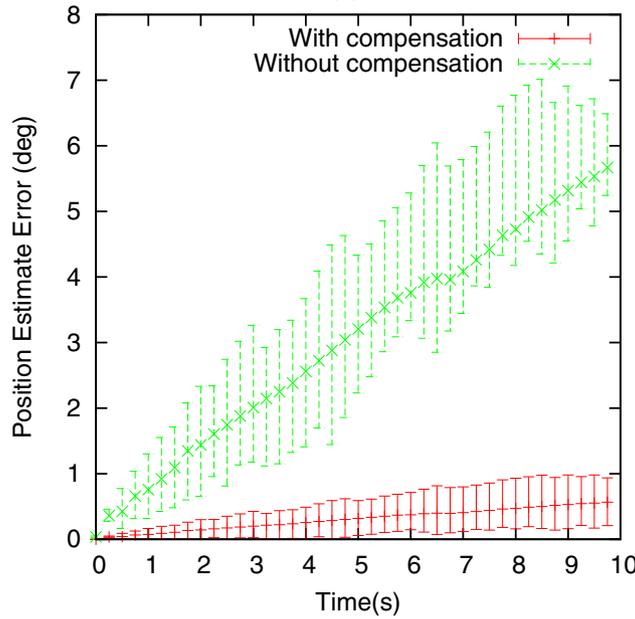
The Ouchi illusion (Ouchi, 1977; Spillmann & Werner, 2012) is perhaps the best known example of an illusory motion effect which can be induced by fixational eye movements. The image consists of centre and surround regions containing two orthogonally oriented rectangular chequerboard patterns (see Fig. 1). The illusion takes the form of a perceived sliding or jittering of the central region with respect to the background. Numerous studies have characterised the illusory motion as resulting from the aperture problem of local motion estimation (Ashida, Kitaoka, & Sakurai, 2005; Fermüller, Pless, & Aloimonos, 2000; Hine, Cook, & Rogers, 1997), created by the juxtaposition of grating patterns of differing orientation. However, while such models can correctly predict the direction of perceived motion in the Ouchi illusion, they provided little insight into why such motion signals are not suppressed when resulting from FEM. Furthermore, the closely related "out of focus" illusion produces a similar illusory motion percept without containing any such orientation structure (Kitaoka, 2001). This illusion consists of a low frequency sinusoidal grating pattern surrounded by a square-wave chequerboard, and can be considered as a generalisation of the Ouchi illusion, with a degree of rotational symmetry. The effects of these two stimuli constitute failures of the mechanism for motion cancellation with respect to fixational eye movements, and as such, a valid model for invariant image processing in the human visual system should also reproduce these effects. We show that these effects arise in our model, resulting from the spatial-frequency contrast at the border between centre and surround regions.

2. Methods and models

We employ here a simple model of RGC integration, in which both the linear response of midget cells and the non-linear response of parasol cells can arise from a centre-surround spatial filtering, as shown in Fig. 2. The model assumes a one-to-one correspondence between photoreceptors, bipolar cells and midget ganglion cells, which is typical of the primate fovea. Thus the width



(a)



(b)

Fig. 8. Top: estimate of object position from retinal motion compared to estimate from invariant motion detector. **Bottom:** position estimation errors for motion detectors, with and without cancellation of global motion.

of the midget cell RF centre is determined by the photoreceptor spacing, which is fixed at a value of 0.5 arcmin.

By contrast, parasol receptive fields are divided into independent sub-fields, corresponding to individual bipolar cells within the RF of the cell (Lee et al., 1998). The spatial components of these sub-field RFs, when combined, give the standard centre-surround structure. However, when convolved with the stimulus the signal contributed by each sub-field is independently rectified before summation, giving a strictly non-negative contribution from each sub-field. This model can qualitatively reproduce many of the complex non-linear behaviours characteristic of parasol or Y-type cells in mammalian retina.

2.1. Linear and non-linear RGCs

In keeping with classical models of the retina, we define the spatial component of the RF as a Difference of Gaussians (DoG) (Wilson & Giese, 1977):

$$\Phi(\mathbf{x}) = \exp\left(-\frac{(\mathbf{x}-\mathbf{x}_0)^2}{2\sigma^2}\right) - \alpha \exp\left(-\frac{(\mathbf{x}-\mathbf{x}_0)^2}{2\beta^2\sigma^2}\right) \quad (1)$$

where the centre to surround ratios of height and width are controlled by the scaling factors α and β .

This filter, which is commonly used in image processing as an edge-enhancing or redundancy reduction filter, has an intuitive

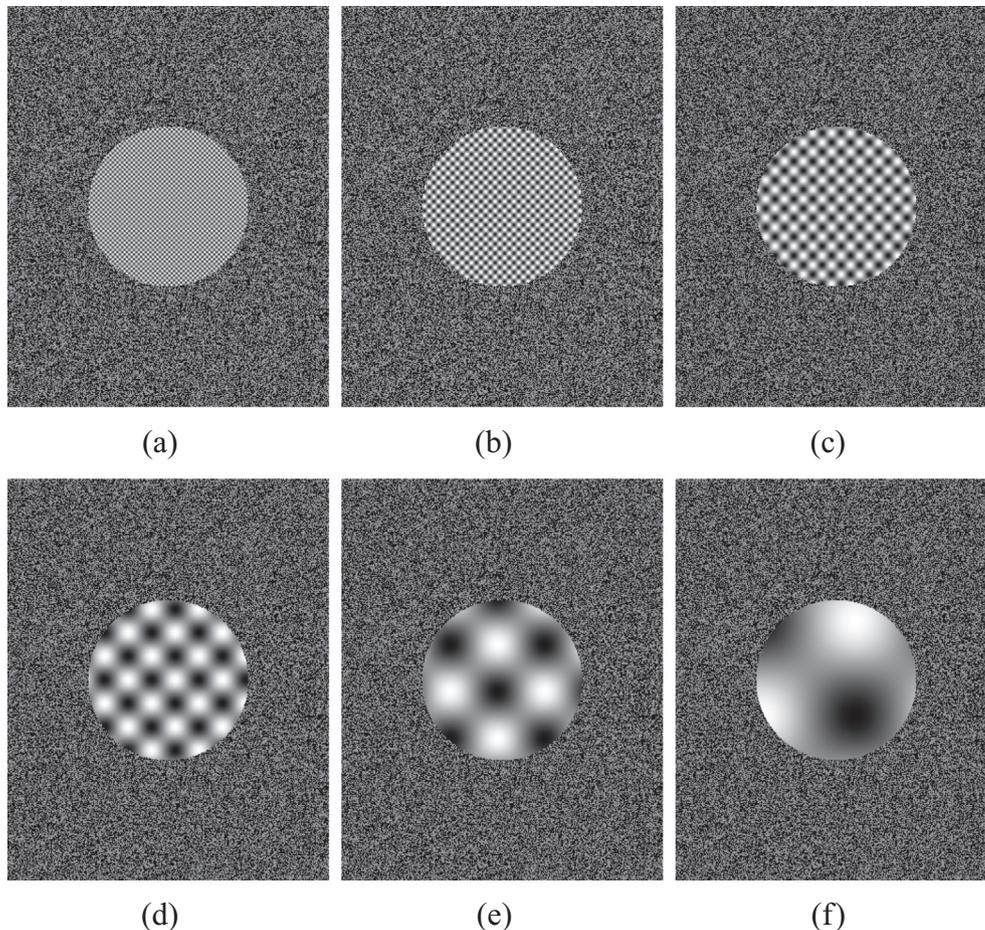


Fig. 9. Example stimuli used in simulation of “Out of focus” illusion. Centre region pattern is generated as the outer product of orthogonally oriented sinusoids, with matching spatial frequencies. Surround is a random chequer pattern with check size of 3 arcmin, corresponding to a fundamental frequency of 10 cyc/deg. Centre spatial frequencies were varied in the range from 0.5–10 cyc/deg.

interpretation in the retina as the summation of inputs from opponent excitatory and inhibitory channels, mediated by the underlying bipolar and amacrine cell circuitry.

Similarly, the temporal component of the receptive field is described by the difference of two independent temporal filters. Each of these two filters is itself the sum of two transient components of opposite sign. This results in a typical bi-phasic filter shape, defined by the time constants τ_1 and τ_2 and scaling parameter p , which can be interpreted as a temporal contrast enhancer (Pillow et al., 2008).

$$f(t) = \frac{t^3}{\tau_1^4} \exp^{-t/\tau_1} - p \frac{t^3}{\tau_2^4} \exp^{-t/\tau_2} \quad (2)$$

By associating each of these two temporal filters with one component of the spatial DoG filter, we obtain a rank-2 spatio-temporal filter, in which each spatial component is modulated by an independent temporal component. Thus, mathematically the RF has the form of a sum of two outer products.

Assuming a locally constant Receptive Field size imposed by the photoreceptor spacing of 0.5 arcmin, this allows us to describe the RF using only eight parameters: two spatial scaling parameters, which define the relative height and standard deviation of the Gaussian components, and the six temporal parameters which define the two bi-phasic filters. The resulting filter can then be convolved with the stimulus and the output rectified to give a prediction of the membrane potential or firing rate for linear midget cells.

Our model of parasol-type cells as shown in Fig. 3 involves a modification to the linear model, in which the RF is divided into

independent sub-fields, which are individually rectified (Hochstein & Shapley, 1976b; Takeshita & Collisch, 2014; Victor et al., 1977). While foveal midget cells generally have only one bipolar cell in their receptive field centre, parasol cells are known to have significantly larger RFs, and thus receive input from many more bipolar cells (Dacey & Petersen, 1992). Hence, each photoreceptor-bipolar cell pairing in the receptive field can be considered as an independent sub-unit, which implements its own rectifying non-linearity. Thus the activity of each sub-unit is the rectified weighted sum of two bi-phasic temporal components, where the weighting of these components is determined from the Gaussians of the parasol cell spatial RF. The RF centre and surround are set to be 4 times larger than those of our linear cells, giving an RF centre width of 2 arcmin.

This relatively simple model can explain a number of qualitative effects observed in mammalian retina.

2.2. Harmonic response to sinusoidal gratings

RGCs are commonly characterised by their response to periodically varying stimuli. When driven by a sinusoidally reversed grating, the firing rate of linear RGCs is phase-locked to the reversal frequency of the grating. This behaviour is easily captured by the linear RF model. By contrast, many non-linear cell types, including parasol cells, respond not just at the fundamental frequency, but, for gratings of high spatial frequency, also display a strong response at the second harmonic of the reversal frequency (Crook et al., 2008; Hochstein & Shapley, 1976a; Shapley & Perry, 1986).

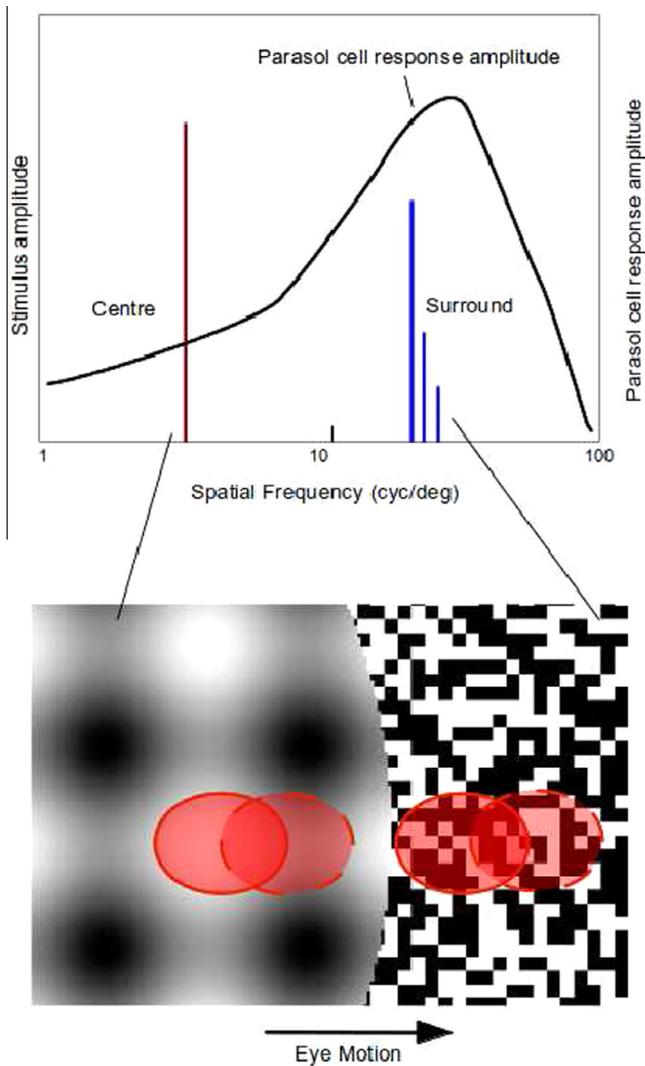


Fig. 10. Illusory motion in the “out of focus” illusion. High spatial frequencies in the surround drive parasol cell responses under FEM, while low spatial frequency centre does not. As a result the motion signal is non-global in the region of the boundary, producing a net motion signal.

This phenomenon, first observed in the Y-Cell of the cat retina (Hochstein & Shapley, 1976b), is well captured by our non-linear model.

Fig. 4 shows the fundamental and second harmonic components of the firing rate of our model cells as a function of the spatial frequency of the stimulus gratings. Midget cells show no second harmonic response. Parasol cells respond with significant components at both fundamental and second harmonic frequencies. For high spatial frequencies, the second harmonic dominates. Notably, the parasol second harmonic curve closely resembles the first harmonic response seen in midget cells. In addition, the peaks of the two parasol cell curves occur at spatial frequencies differing by a factor of 4, corresponding to the scaling between parasol cell and sub-field RF.

This suggests an intuitive interpretation of this effect, whereby stimuli containing high spatial frequencies independently excite sub-fields within the parasol cell RF, leading to a frequency doubled response.

2.3. Phase invariance and response to saccades

The intuitive interpretation of the centre-surround RGC filter suggests that centre-surround cells should respond most strongly

to regions of high spatial contrast or edges within the stimulus image. This is found to be generally true of midget cells, and when stimulated with periodic gratings, the response of midget cells is found to depend strongly on the relative phase of the grating with respect to the RF. The same is not true of parasol cells, which are often observed to respond strongly to such gratings irrespective of the stimulus phase (Crook et al., 2008). This phase invariant property is reproduced by our model parasol cells, and leads to a characteristic population response to simulated microsaccades, as shown in Fig. 5.

This characteristic of parasol cells is easily explained as a consequence of the non-linear integration performed in our model. While linear RGCs respond to both spatial and temporal contrast across the entire RF, the individual sub-fields on the parasol RF respond only to local temporal contrast. Since sub-field contributions are rectified before summation, centre and surround activations do not cancel out. Thus parasol cells can respond strongly to time varying stimuli, even in the absence of spatial contrast.

3. Motion detection and cancellation circuits

Motion detection is a primary task of the visual system, and numerous studies have found motion-sensitive neurons at all levels of the early visual system. However, whereas direction-selective cells are commonly found in the retinae of amphibians and some small mammals (Barlow, Hill, & Levick, 1964; Öveczky, Baccus, & Meister, 2003; Vaney, He, Taylor, & Levick, 2001), they are rarely reported in the primate or human retina. Thus, here we assume that the estimation of motion direction and velocity occurs not in the retina itself, but in higher visual areas receiving topographic input from the retina.

The Reichardt-Hassenstein detector (Hassenstein & Reichardt, 1956; Reichardt, 1961) is a well known model of neural motion detection. The detector consists of a small local network in which an output cell detects a temporally delayed correlation in the activity of two nearby ganglion cells. Thus the detector is selective to local motion with a certain velocity and direction, determined by the delay and relative position of the ganglion cells respectively. Here the local motion on the retina is computed by performing a vector sum over the activity of a local population of such detectors. The motion vector contributed by each detector is determined by its velocity and direction tuning, with its weight given by the instantaneous firing rate. A detailed description of the Reichardt-Hassenstein detector and related methods of motion detection can be found in Reichardt (1987). However, for our purposes the details of the motion detection circuitry are not important. It is sufficient to assume a mechanism whereby a local motion vector may be computed within each region of the visual field based purely on retinal motion. The task then remains to distinguish true motion signals from false alarms caused by FEM.

The strong transient response of parasol cells to simulated eye-movements suggests a simple retinal mechanism whereby local image motion can be distinguished from apparent global motion induced by eye movements. In this system, illustrated in Figs. 6 and 11, the retinal motion signal within a region of the visual field is computed locally by a population of Reichardt-like motion detectors. This signal is then gated by the output of a non-linear thresholding neuron, which receives input from the parasol RGC population. If the threshold of this cell is set such that it fires only in response to a critical number of nearly coincident spikes, its firing will indicate coherent motion within its receptive field (see Fig. 7). We simulate a population of such cells tiling the retina, with a receptive field width of 10 arcmin, corresponding to 5 parasol RF centre widths, where each cell gates the activity of the motion detector circuit within its own RF. As a result, motion signals should be suppressed except where the RF of the threshold neuron

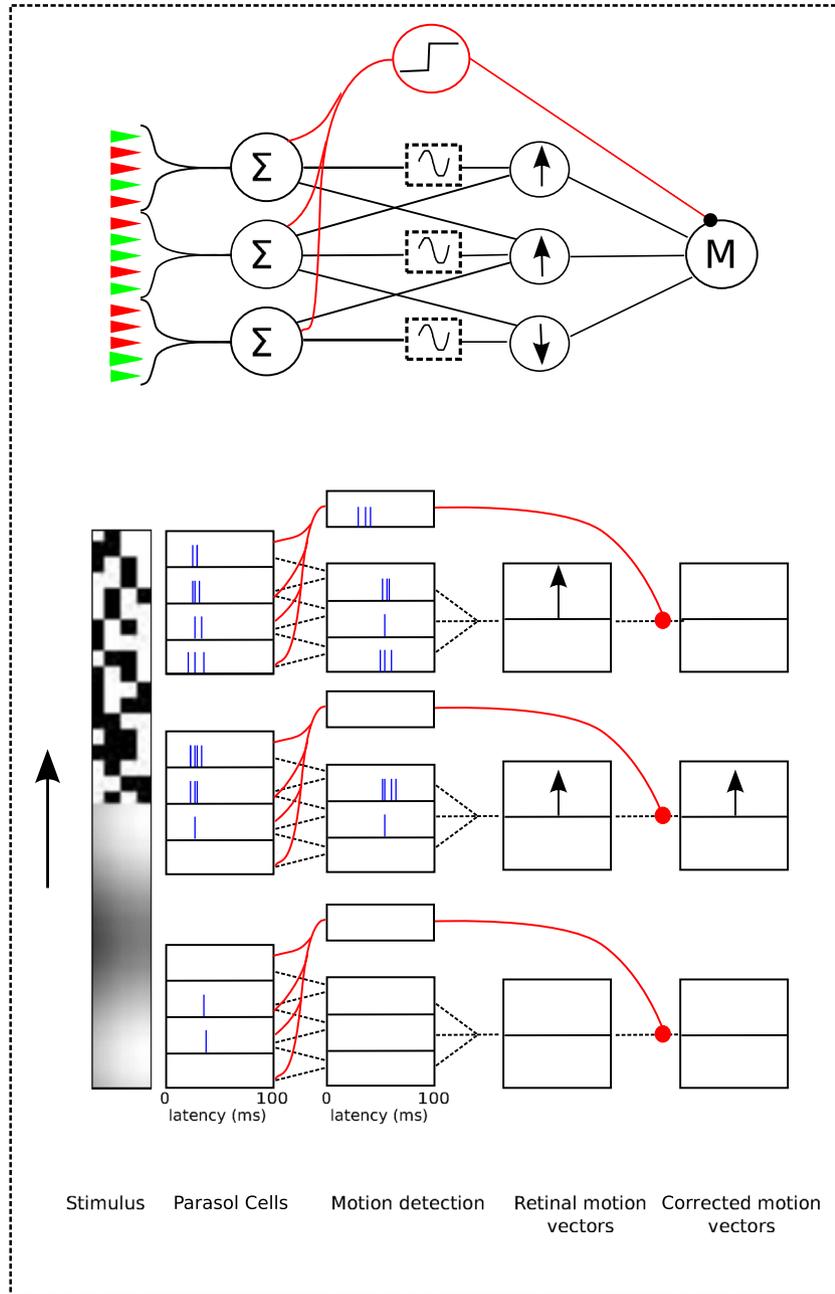


Fig. 11. Top: network model for motion detection and cancellation circuit. **Bottom:** network response to the “Out of focus” image undergoing a single simulated eye movement. The plot shows a small cross section of the activity of three local populations within the retina, each of which feeds forward to a local motion detection layer. The output of the motion detectors is gated by the activity of the global motion threshold unit in each population. Activities of parosol cells and motion detectors are plotted as spike rasters, while computed motion vectors are represented by vertical arrows. For simplicity, only a small cross section of the network is shown, and only the vertical component of motion is plotted. The image undergoes a simulated rapid eye movement of amplitude 5 arcmin in the vertical direction. This motion elicits strong responses in parosol cells whose RFs lie in the high frequency region of the image, but little or no activity in those within the low-frequency region. Motion vectors corresponding to high frequency regions are subsequently suppressed due to the activity of the global motion unit. Only motion vectors generated in the border region are preserved.

lies across the boundary between moving and stationary regions of the visual scene. In the case of uniform motion throughout the visual scene, motion signals will be suppressed entirely.

Thus the local motion vector, $\mathbf{M}(t)$ computed by this circuit can be approximated as

$$\mathbf{M}(t) \simeq \left(\sum_i \mathbf{v}_i a_i(t) \right) \left(1 - \Theta \left(\sum_j r_j(t) \right) \right) \quad (3)$$

where the vectors \mathbf{v}_i and a_i are the motion vectors and activities associated with the Reichardt detectors, r_j are the activities of the

parosol cells, and $\Theta()$ is the threshold function, which equals 1 when the threshold is exceeded, and 0 otherwise.

4. Results

4.1. Invariant motion detector model

As a result of the non-linear summation of their spatial RFs, our model parosol cells display a fast, transient, phase-invariant response to saccadic stimulus motion. These cells respond strongly to stimulus changes within their RF, largely independently of

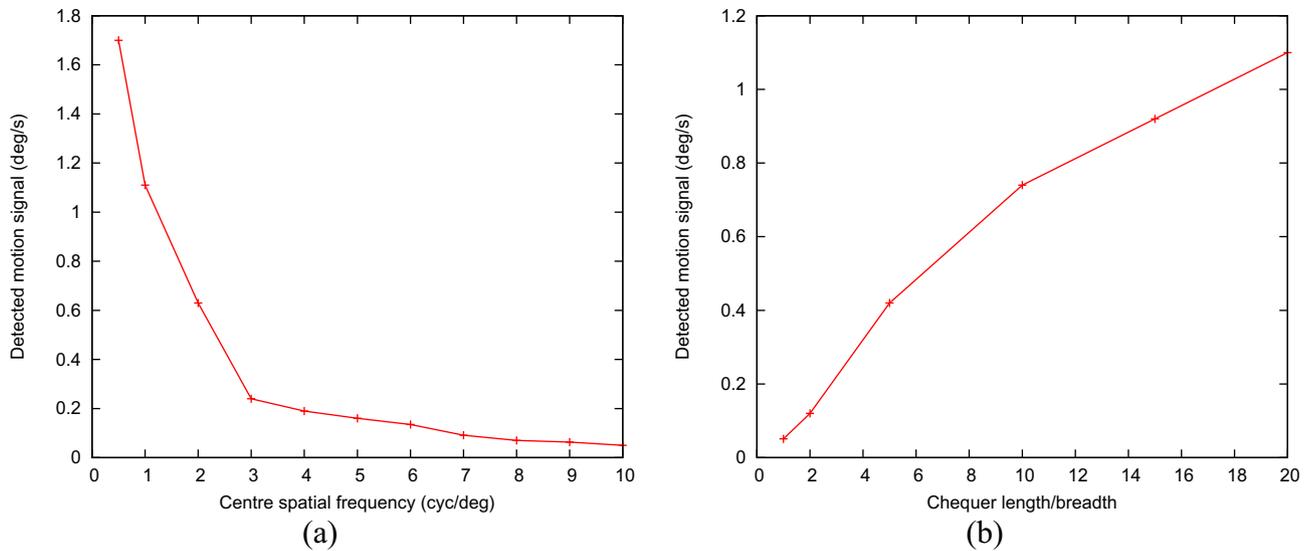


Fig. 12. **Left:** detected motion signal produced by FEM as a function of centre spatial frequency in the “Out of focus” illusion. At lower spatial frequencies the motion cancellation mechanism fails, leaving a net motion percept. When centre spatial frequency matches the fundamental frequency of the surround, the motion signal is almost entirely cancelled out. **Right:** motion signal produced by FEM in the Ouchi illusion. The net motion signal appears at high values of the length-to-breadth ratio, corresponding to higher contrast in spatial frequencies along orthogonal directions.

spatial phase or scale. As a result, the population response of parasol ganglion cells can provide a retinal signal for coherent global motion.

We test this model’s ability to track the position of a visual object moving against a background of natural scenes in the presence of simulated fixational eye movements. The stimulus consists of an object, extracted from a natural scene, moving against a background taken from the same natural image. Image resolution was chosen to give a pixel size five times smaller than the receptive field of a photoreceptor in our model. The entire visual scene, including both target and background underwent continual motion to simulate fixational eye movements, consisting of slow linear drift, Gaussian tremors, and ballistic microsaccades. Saccade amplitude and velocity are drawn at random from the “main sequence” of microsaccades (Bahill, Clark, & Stark, 1975; Otero-Millan, Troncoso, Macknik, Serrano-Pedraza, & Martinez-Conde, 2008), with amplitudes in the range 0.1–1 deg. Given a fixed starting point, real world object position is estimated by integrating the output of the motion detector circuit. Fig. 8(a) below shows the object position as estimated by motion on the retina alone (green), as well as the estimate obtained by suppression of global motion signals (blue). Fig. 8(b) shows the error in position estimates in both cases. Our motion cancellation model gives a much improved estimate of object position.

4.2. Prediction of illusory jitter in the Ouchi and Out-of-focus illusions

An intriguing consequence of the motion-cancellation model proposed here is that the suppression of eye-motion percepts may fail under certain conditions of spatial frequency contrast – the same conditions under which the illusory motion of the Ouchi and Out-of-focus illusions is observed. An implicit assumption in our model is that natural scenes are scale-invariant. That is to say, they contain visual information at all spatial frequencies, according to the familiar $1/f^2$ power spectrum (van der Schaaf & van Hateren, 1996). Thus it is assumed that, following an eye movement, the new target region contains sufficient contrast at the appropriate spatial scale to produce a population response in the non-linear cells. While robust under normal viewing conditions, this mechanism may fail when presented with stimuli

having narrow-band spatial frequency spectra. Specifically, the system may fail when presented with a stimulus in which a region containing little or no high frequency content directly borders a region of high spatial frequencies.

This effect is illustrated in Fig. 10: at an appropriate spatial scale the high frequencies of the square-wave surround in the “out of focus” illusion strongly drive the second harmonic response of the non-linear cells, while the low frequency centre does not. Hence, eye movements which trigger a motion response in ganglion cells in the high frequency region may not produce a response in neighbouring cells whose RFs lie across the boundary in the low frequency region. Thus a global motion may produce a purely local motion signal in the region of the boundary.

In the case of the Ouchi illusion, the chequerboard pattern contains high spatial frequencies in one direction, and low spatial frequencies in the orthogonal direction, leading to an abrupt change in spatial frequency when crossing the centre-surround boundary in either the vertical or horizontal direction. In the case of the out-of-focus illusion, this transition from low to high spatial frequency occurs at all points on the boundary, independent of orientation.

We test our model’s response to these images under simulated FEM as a function of the spatial frequency (see Fig. 9). In the Ouchi illusion, the spatial frequency is varied by changing the ratio of length to breadth in the chequer, with a value of 1 corresponding to a square grid. In both cases, the high frequency square wave component in the surround is fixed to a fundamental frequency of 10 cyc/deg. Simulated fixational eye movements consist of microsaccades, with mean inter-saccadic interval of 300 ms, and mean amplitude of 0.5 deg, combined with slow drift (velocity 0.3 deg/s) and Gaussian tremor, to give a mean integrated motion of 2 deg/s.

Fig. 12 shows the net motion signal produced by these images as a function of centre spatial frequency. The net motion signal is computed by integrating the motion detector output over the duration of the simulation, with a value of 2 deg/s corresponding to complete failure of the cancellation mechanism, and 0 deg/s to a complete cancellation of the retinal motion. In both cases our model predicts a strong net motion signal under FEM for spatial frequencies in the range from 0.2–3 cyc/deg. This agrees well with

the psychophysical results of [Spillmann \(2013\)](#) and [Khang and Essock \(1996\)](#), who report that the Ouchi illusion is strongest for frequencies of 0.5–2.5 cyc/deg, and 1–1.5 cyc/deg respectively. The motion signal is progressively suppressed at higher spatial frequencies, leading to near complete cancellation where the centre frequency matches the surround frequency of 10 cyc/deg.

5. Discussion

The questions of whether and how retinal motion signals can be corrected for errors due to FEM have been discussed for some time. While early theories suggested that stabilisation under FEM was either unnecessary, or dependant on explicit eye-motion information, in the form of efference copy or corollary discharge signals, recent studies have supplied compelling evidence that motion correction with respect to FEM is both necessary and a result of purely retinal mechanisms. In addition to the many examples of illusory motion effects consistent with purely retinal mechanisms ([Murakami & Cavanagh, 1998](#); [Murakami et al., 2006](#); [Spillmann, 2013](#)), several experiments have ruled out extra-retinal correction, and provided strong evidence that the human visual system primarily encodes differential motion, rather than retinal motion: [Murakami \(2004\)](#) found a correlation between fixational instability and motion detection thresholds, while also demonstrating that thresholds are lowered by the presence of a visual surround, providing a reference frame for differential motion. Subsequently, [Poletti et al. \(2010\)](#) performed an elegant experiment in which the visual scene was stabilised on the retina using a gaze-contingent display. They demonstrated conclusively that motion detection is dependent on differential motion within the scene, and not on absolute retinal motion. In particular, they showed that a moving stimulus may appear stationary, even while being tracked by eye movements.

The nature of the retinal motion cancellation, however, remains unclear. One candidate mechanism is suggested by [Ölveczky et al. \(2003\)](#) and [Baccus et al. \(2008\)](#) who have identified object motion sensitive (OMS) ganglion cells in the retinae of salamander and rabbit which are selective for relative motion between RF centre and surround. Unlike the rabbit, however, the primate retina contains relatively few motion sensitive cells ([Bach & Hoffmann, 2000](#)), making such a mechanism seem impractical in human vision. Indeed, given the relative lack of specialisation in primate RGCs, and the observation that human motion processing takes place at the level of the cortex, we suggest that mechanisms which make use of established retinal pathways are more viable candidates for human motion cancellation. In addition, [Spillmann \(2013\)](#) notes that the spatial frequency dependence of the Ouchi illusion, and other related motion effects suggests a magnocellular origin for such a mechanism.

The model we have presented here employs a principle similar to that of the OMS cells in salamander retina, using local motion contrast to distinguish object motion from retinal motion. However, our model relies solely on properties of magnocellular-projecting RGCs which are ubiquitous in primate retina, and does not require the existence of specialised retinal circuitry which has yet to be identified in primates.

A consequence of this difference is that both motion detection and cancellation in our model occur at later stages in the visual pathway. For example, FEM may drive increased activity in early cortical areas such as V1, which is subsequently suppressed in higher motion-specific areas. As a result FEM may be expected to elicit enhancement of activity in direction selective cells in primary cortex, while simultaneously leading to suppression in downstream motion sensitive cells. This is in keeping with the results of [Herrington et al. \(2009\)](#), who reported that microsaccades sup-

press activity in macaque area MT and other dorsal stream areas during motion detection tasks. Indeed, [Martinez-Conde et al.](#) offer a summary of the effects of microsaccades on cortical activity ([Martinez-Conde et al., 2013](#)), which demonstrates how FEM may induce suppression of neural activity in the motion-processing dorsal stream ([Bair & O'Keefe, 1998](#); [Herrington et al., 2009](#)) while at the same time leading often to enhancement in V1, V2 and ventral areas. This view is consistent with that proposed in our model, where activity is increased in the first layer of the motion detector circuit, but suppressed in the higher FEM-invariant motion-sensitive cells, which may be considered roughly analogous to MT or higher dorsal stream areas.

Furthermore, unlike the proposal of [Murakami and Cavanagh](#), our model does not require any complicated lateral interactions, or any global velocity estimate in order to cancel spurious motion signals. Indeed, while it is possible to imagine any number of mechanisms which might achieve motion correction through lateral connections at the cortical level, perhaps taking advantage of the phase invariance of complex cells, it is worth noting that the information necessary to weed out erroneous motion signals is already present in the earliest stages of visual processing. This is shown explicitly in [Fig. 11](#), where both local and global motion is computed directly from retinal signals in the second layer of the network, which receives direct input from the magnocellular-projecting retinal pathway, thus allowing the subsequent cortical layers of the network to achieve gaze-invariance. A distinct advantage of the model proposed here over possible models involving cortical interactions is that in this case the cancellation mechanism arises “for free” as a result of the spatial frequency response properties of parasol cells without the requirement of additional circuitry.

It must, however, be noted that this model does not describe a complete framework for invariant motion processing. While our model allows for the generation of purely local motion vectors and thus for the identification of local motion boundaries, complex processes of object separation and binding would also be required in order to correctly determine what is moving relative to what. The problem of figure-ground segregation is the subject of active research fields in its own right, with several candidate mechanisms including coding of border ownership through lateral interactions in V2 ([Zhou, Friedman, & von der Heydt, 2000](#); [Zwicker, Wachtler, & Eckhorn, 2007](#)), while the perennial “Binding Problem” ([von der Malsburg, 1994](#)) has been a central theme of visual research since it was first posed in the 1980s. We make no claim to address such questions here. Instead, we have simply assumed that such mechanisms, since they seem necessary, must exist, and confined ourselves to a discussion of how motion signals may be computed and erroneous signals suppressed based on retinal information. As such, our results apply only to the processing of motion signals in general, rather than to the motion of specific objects.

Despite this limitation, and in contrast to other published hypotheses, our model offers an insight into two commonly observed visual illusions. The Ouchi illusion has been widely discussed as an example of the aperture problem in local motion estimation, resulting from the presence of oriented gratings. This is perhaps a result of the fact that the illusion is often observed following directed eye movements, which bias the direction of expected retinal motion. However, it is often overlooked that the illusion can also be observed, albeit less strongly, during attempted fixation. Furthermore, the “out of focus” illusion produces a similar, and in fact stronger motion effect during fixation, without containing the orthogonally oriented gratings of the Ouchi illusion. Indeed, it seems likely that there are in fact two effects visible in the Ouchi illusion; one, triggered by directed eye movements which results from the aperture problem, and a second, triggered

by fixational eye movements, resulting from the failure of the motion cancellation circuit. This second effect is best exemplified by the “out of focus” image, which in this context can be considered a rotation invariant version of the Ouchi image, and thus produces a stronger effect, since the illusory motion is largely independent of the direction of eye motion. It is instructive to note that psychophysical studies of the Ouchi illusion have reported a strong spatial frequency dependence matching that predicted by our model (Khang & Essock, 1996; Spillmann, 2013), and corresponding closely to the RF width of foveal parasol ganglion cells. This observation further supports the suggestion that the mechanism of motion cancellation is of retinal, rather than cortical origin.

A further interesting consequence of our model of motion encoding is the prediction that, in the absence of any fixed reference point or other sensory input, a uniform and coherent motion of the entire visual scene should not produce any motion percept. While this may seem surprising at first, it is important to note that such purely uniform motion rarely if ever occurs under natural viewing conditions, and where it does occur, it is usually accompanied by vestibular motion information.

Acknowledgment

This work was supported by the Bernstein Center for Computational Neuroscience Munich (BMBF Grant 01GQ1004A).

References

- Ahissar, E., & Arieli, A. (2012). Seeing via miniature eye movements: a dynamic hypothesis for vision. *Frontiers in Computational Neuroscience*, 6.
- Ashida, H., Kitaoka, A., & Sakurai, K. (2005). A new variant of the Ouchi illusion reveals fourier-component-based processing. *Perception*, 34, 381–390.
- Baccus, S. A., Ölveczky, B. P., Manu, M., & Meister, M. (2008). A retinal circuit that computes object motion. *The Journal of Neuroscience*, 28(27), 6807–6817.
- Bach, M., & Hoffmann, M. B. (2000). Visual motion detection in man is governed by non-retinal mechanisms. *Vision Research*, 40(18), 2379–2385.
- Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, 24(3), 191–204.
- Bair, W., & O’Keefe, L. P. (1998). The influence of fixational eye movements on the response of neurons in area MT of the macaque. *Visual Neuroscience*, 15(04), 779–786.
- Barlow, H., Hill, R., & Levick, W. (1964). Retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit. *The Journal of Physiology*, 173(3), 377–407.
- Bölinger, D., & Gollisch, T. (2012). Closed-loop measurements of iso-response stimuli reveal dynamic nonlinear stimulus integration in the retina. *Neuron*, 73(2), 333–346.
- Burak, Y., Rokni, U., Meister, M., & Sompolinsky, H. (2010). Bayesian model of dynamic image stabilization in the visual system. *Proceedings of the National Academy of Sciences*, 107(45), 19525–19530.
- Crook, J. D., Peterson, B. B., Packer, O. S., Robinson, F. R., Troy, J. B., & Dacey, D. M. (2008). Y-cell receptive field and collicular projection of parasol ganglion cells in macaque monkey retina. *The Journal of Neuroscience*, 28(44), 11277–11291.
- Dacey, D. M., & Petersen, M. R. (1992). Dendritic field size and morphology of midsize and parasol ganglion cells of the human retina. *Proceedings of the National Academy of Sciences*, 89(20), 9666–9670.
- Enroth-Cugell, C., & Robson, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *The Journal of Physiology*, 187(3), 517–552.
- Enroth-Cugell, C., Robson, J., Schweitzer-Tong, D., & Watson, A. (1983). Spatio-temporal interactions in cat retinal ganglion cells showing linear spatial summation. *The Journal of Physiology*, 341(1), 279–307.
- Fermüller, C., Pless, R., & Aloimonos, Y. (2000). The Ouchi illusion as an artifact of biased flow estimation. *Vision Research*, 40(1), 77–95.
- Gollisch, T., & Meister, M. (2010). Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron*, 65(2), 150–164.
- Gregory, R. (1958). Eye movements and the stability of the visual world. *Nature*, 182(4644), 1214.
- Hafed, Z. M. (2011). Mechanisms for generating and compensating for the smallest possible saccades. *European Journal of Neuroscience*, 33(11), 2101–2113.
- Hafed, Z. M., & Krauzlis, R. J. (2010). Microsaccadic suppression of visual bursts in the primate superior colliculus. *The Journal of Neuroscience*, 30(28), 9542–9547.
- Hassenstein, B., & Reichardt, W. (1956). Systemtheoretische analyse der zeit-, reihenfolgen- und vorzeichenbewertung bei der bewegungsperzeption des räusselfäfers chlorophanus. *Zeitschrift für Naturforschung*, 11, 513–524.
- Herrington, T. M., Masse, N. Y., Hachmeh, K. J., Smith, J. E., Assad, J. A., & Cook, E. P. (2009). The effect of microsaccades on the correlation between neural activity and behavior in middle temporal, ventral intraparietal, and lateral intraparietal areas. *The Journal of Neuroscience*, 29(18), 5793–5805.
- Hine, T., Cook, M., & Rogers, G. T. (1997). The Ouchi illusion: An anomaly in the perception of rigid motion for limited spatial frequencies and angles. *Perception & Psychophysics*, 59(3), 448–455.
- Hochstein, S., & Shapley, R. (1976a). Linear and nonlinear spatial subunits in y cat retinal ganglion cells. *The Journal of Physiology*, 262(2), 265–284.
- Hochstein, S., & Shapley, R. (1976b). Quantitative analysis of retinal ganglion cell classifications. *The Journal of Physiology*, 262(2), 237–264.
- Khang, B.-G., & Essock, E. A. (1996). Apparent relative motion from a checkerboard surround. *Perception*, 26(7), 831–846.
- Kitaoka, A. (2001). Out of focus.
- Kremers, J., Lee, B. B., Pokorny, J., & Smith, V. C. (1993). Responses of macaque ganglion cells and human observers to compound periodic waveforms. *Vision Research*, 33(14), 1997–2011.
- Kuffler, S. W. et al. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16(1), 37–68.
- Lee, B. B., Kremers, J., & Yeh, T. (1998). Receptive fields of primate retinal ganglion cells studied with a novel technique. *Visual Neuroscience*, 15(01), 161–175.
- Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 207(1167), 187–217.
- Martinez-Conde, S. (2006). Fixational eye movements in normal and pathological vision. *Progress in Brain Research*, 154, 151–176.
- Martinez-Conde, S., Macknik, S. L., & Hubel, D. H. (2000). Microsaccadic eye movements and firing of single cells in the striate cortex of macaque monkeys. *Nature Neuroscience*, 3(3), 251–258.
- Martinez-Conde, S., Otero-Millan, J., & Macknik, S. (2013). The impact of microsaccades on vision: Towards a unified theory of saccadic function. *Nature Reviews Neuroscience*, 14(2), 83–96.
- Meirovitz, E., Ayzenshtat, I., Werner-Reiss, U., Shamir, I., & Slovov, H. (2011). Spatiotemporal effects of microsaccades on population activity in the visual cortex of monkeys during fixation. *Cerebral Cortex*, page bhr102.
- Münch, T. A., Da Silveira, R. A., Siebert, S., Viney, T. J., Awatramani, G. B., & Roska, B. (2009). Approach sensitivity in the retina processed by a multifunctional neural circuit. *Nature Neuroscience*, 12(10), 1308–1316.
- Murakami, I. (2004). Correlations between fixation stability and visual motion sensitivity. *Vision Research*, 44(8), 751–761.
- Murakami, I., & Cavanagh, P. (1998). A jitter after-effect reveals motion-based stabilization of vision. *Nature*, 395(6704), 798–801.
- Murakami, I., Kitaoka, A., & Ashida, H. (2006). A positive correlation between fixation instability and the strength of illusory motion in a static display. *Vision Research*, 46(15), 2421–2431.
- Ölveczky, B. P., Baccus, S. A., & Meister, M. (2003). Segregation of object and background motion in the retina. *Nature*, 423(6938), 401–408.
- Otero-Millan, J., Macknik, S. L., & Martinez-Conde, S. (2012). Microsaccades and blinks trigger illusory rotation in the rotating snakes illusion. *The Journal of Neuroscience*, 32(17), 6043–6051.
- Otero-Millan, J., Troncoso, X. G., Macknik, S. L., Serrano-Pedraza, I., & Martinez-Conde, S. (2008). Saccades and microsaccades during visual fixation, exploration, and search: foundations for a common saccadic generator. *Journal of Vision*, 8(14), 21.
- Ouchi, H. (1977). *Japanese optical and geometrical art*. Dover.
- Perrone, J. A., & Krauzlis, R. J. (2008). Vector subtraction using visual and extraretinal motion signals: A new look at efference copy and corollary discharge theories. *Journal of Vision*, 8(14), 24.
- Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E., & Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207), 995–999.
- Pitkow, X., Sompolinsky, H., & Meister, M. (2007). A neural computation for visual acuity in the presence of eye movements. *PLoS Biology*, 5(12), e331.
- Poletti, M., Listorti, C., & Rucci, M. (2010). Stability of the visual world during eye drift. *The Journal of Neuroscience*, 30(33), 11143–11150.
- Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. *Sensory Communication*, 303–317.
- Reichardt, W. (1987). Evaluation of optical motion information by movement detectors. *Journal of Comparative Physiology A*, 161(4), 533–547.
- Rodieck, R. W. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 5(12), 583–601.
- Shapley, R., & Perry, V. (1986). Cat and monkey retinal ganglion cells and their visual functional roles. *Trends in Neurosciences*, 9, 229–235.
- Spillmann, L. (2013). The Ouchi–Spillmann illusion revisited. *Perception*, 42, 413–429.
- Spillmann, L., & Werner, J. S. (2012). *Visual perception: The neurophysiological foundations*. Elsevier.
- Steinman, R. M., Haddad, G. M., Skavenski, A. A., & Wyman, D. (1973). Miniature eye movement. *Science*, 181(4102), 810–819.
- Takeshita, D., & Gollisch, T. (2014). Nonlinear spatial integration in the receptive field surround of retinal ganglion cells. *The Journal of Neuroscience*, 34(22), 7548–7561.
- Tong, J., Lien, T. C., Cisarik, P. M., & Bedell, H. E. (2008). Motion sensitivity during fixation in straight-ahead and lateral eccentric gaze. *Experimental Brain Research*, 190(2), 189–200.
- van der Schaaf, A., & van Hateren, J. (1996). Modelling the power spectra of natural images: Statistics and information. *Vision Research*, 36(17), 2759–2770.
- Vaney, D. I., He, S., Taylor, W. R., & Levick, W. R. (2001). Direction-selective ganglion cells in the retina. In *Motion vision* (pp. 14–57). Springer.

- Victor, J. D., & Shapley, R. M. (1979). Receptive field mechanisms of cat x and y retinal ganglion cells. *The Journal of General Physiology*, 74(2), 275–298.
- Victor, J. D., Shapley, R. M., & Knight, B. W. (1977). Nonlinear analysis of cat retinal ganglion cells in the frequency domain. *Proceedings of the National Academy of Sciences*, 74(7), 3068–3072.
- von der Malsburg, C. (1994). *The correlation theory of brain function*. Springer.
- Wilson, H. R., & Giese, S. C. (1977). Threshold visibility of frequency gradient patterns. *Vision Research*, 17(10), 1177–1190.
- Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *The Journal of Neuroscience*, 20(17), 6594–6611.
- Zwicker, T., Wachtler, T., & Eckhorn, R. (2007). Coding the presence of visual objects in a recurrent neural network of visual cortex. *Biosystems*, 89(1–3), 216–226.